Development of Trip Production Models Using Household Surveys for Urban Travel Demand Forecasting

Safa Ali Lafta
Department of Laboratories, Al-Iraqia University, Baghdad, Iraq
safa.a.lafta@aliraqia.edu.iq (corresponding author)

Hiba M. Raheem
Department of Laboratories, Al-Iraqia University, Baghdad, Iraq
Hiba.m.raheem@aliraqia.edu.iq

Anas Nahidh Hassooni
Civil Engineering Department, Al-Iraqia University, Baghdad, Iraq
anasnahidh@aliraqia.edu.iq

Yerevan Adil Ali
Department of Laboratories, Al-Iraqia University, Baghdad, Iraq
yerevan.a.ali@aliraqia.edu.iq

Received: 22 April 2025 | Revised: 8 June 2025, 3 July 2025, 13 July 2025, and 17 July 2025 | Accepted: 19 July 2025

 DOI: https://doi.org/10.48084/etasr.11671

ABSTRACT

Transportation planning is a critical component of the urban development, relying heavily on accurate travel demand forecasting. With the number of vehicles more than doubling, increasing the reliance on private cars has exacerbated the traffic congestion and environmental pollution. This study aims to develop a statistical forecasting model to examine the relationship between the socioeconomic factors and trip generation, thereby supporting data-driven infrastructure improvements to alleviate congestion. Two data collection methods were employed: in-home interviews and questionnaire-based household surveys. The study found that the trip production is significantly influenced by variables, such as the number of students and workers in the household, family size and composition, gender distribution, household age groups, income levels, and car ownership. A total of 5,529 trips were recorded. Trip generation models were developed for three primary journey purposes: home-based work, education, and total trips, using SPSS software. The multiple regression analysis yielded a coefficient of determination (R^2) of 0.850 and an average accuracy (AA%) of 72.5%. The results underscore that the household socioeconomic characteristics are key determinants of the travel behavior, and their integration into planning models can enhance the urban mobility strategies.

Keywords-transportation planning; trip generation; travel demand forecasting; multiple regression model; trips

I. INTRODUCTION

Transport planning is a preparatory process aimed at enabling the efficient movement of people between locations. It constitutes a core component of urban planning and requires a systematic, data-driven approach. Effective transport planning facilitates the integration of multiple dynamic aspects of the urban environment and directly influences three key components of the transportation systems: highway operations, geometric design parameters, and traffic facility management [1-4]. A fundamental element of transportation planning is the travel demand forecasting, which involves a sequential modeling process wherein the output of each phase serves as the input for the next [5]. This modeling is typically grounded in land-use characteristics and the average trip behavior [3]. The travel demand forecasting framework generally comprises the following stages: land use and travel behavior analysis, trip generation, trip distribution, mode choice analysis, and traffic assignment [6-9].

This paper focuses on the trip generation, offering an overview to highlight its role within the broader context of transportation planning [10]. Several studies have contributed to model development and demonstrated the practical applications in transport planning, both regionally and internationally. Authors in [11-13], evaluated the core principles, benefits, and challenges of implementing smart transportation systems in Saudi Arabia, aligning their development with broader economic and technological strategies. Also they emphasized the role of green mobility in supporting Saudi Arabia's Vision 2030, promoting sustainable urban development through data-driven policies aimed at reducing the environmental impact. The impact of Transit-Oriented Development (TOD) in Riyadh has been investigated, stressing its potential to balance the economic growth, environmental sustainability, and social equity. Authors in [2], developed predictive models for trip attraction in the Bab al-Muadham area of Baghdad. Their analysis showed strong correlations between the trip volumes and variables, such as employment density, healthcare capacity, and student populations. Moreover, in [3], the trip generation modeling was conducted in the Hayy Ur area of Baghdad. The study, based on detailed household surveys across 21 traffic zones, revealed that the household composition and job locations significantly influence the travel behavior. The aim of the current study is to identify the primary factors influencing the trip generation and develop statistical models that simulate the travel behavior using household survey data, land use patterns, and socioeconomic variables.

II. PROBLEM DEFINITION

The transportation network is a vital component of the urban infrastructure, reflecting a region's socioeconomic development. In Baghdad, the rapid urbanization and population growth have presented significant transportation challenges. According to the Traffic Police Directorate, the number of registered vehicles has more than doubled since 2003, exceeding the capacity of the existing road infrastructure and contributing to increased congestion and environmental degradation [14-16]. This is further complicated by the ongoing land-use changes and limitations in data collection systems [17]. This study aims to develop a statistical forecasting model to examine the relationship between the household and travel behavior variables and their impact on congestion. The goal is to support data-driven infrastructure upgrades and promote improved quality of life through shorter and safer commutes [18-20]. By addressing these challenges in an integrated manner, Baghdad can enhance its transportation system and move toward sustainable urban development [21]. The primary objectives are:

- To analyze the household travel behavior patterns by investigating trip generation characteristics, such as frequency, purpose, and timing, across various household types.
- To identify the key demographic and socioeconomic variables influencing the travel demand in the study area.
- To develop trip production models using Multiple Linear Regression (MLR) based on variables, such as household size, age structure, income level, and car ownership.

III. RESEARCH METHODOLOGY

This study adopts a structured methodological framework to collect and analyze data for trip generation modeling.

A. Data Collection

A dual-mode survey strategy was implemented, comprising:

- Home Interviews (In-Person): Conducted with household members to collect detailed travel behavior data.
- Distributed Questionnaires: Designed to capture the socio-economic indicators, educational status, trip generation motivators, and travel patterns.

This approach ensures the comprehensive collection of household-level data relevant to the trip generation analysis.

B. Study Area Selection

The study area was determined based on the existing transportation facilities and the extent of the urban region. An external cordon line was used to define the study boundary, ensuring that the research captured the predominant travel patterns in the region [16]. This external boundary was subdivided into zones to facilitate a cdetailed analysis and accurate data collection.

C. Zoning of the Study Area

The study area was subdivided into Traffic Analysis Zones (TAZs) to manage and simplify complex travel data (Figure 1). This zoning helps streamline the representation of trips, households, and employment centers [22]. The main objectives of zoning include [23]:

- Spatial Understanding: Enhances the comprehension of the activity distribution and land-use patterns.
- Data Simplification: Facilitates an easier collection and analysis of survey data.
- Computational Efficiency: Reduces the processing time and complexity in modeling.

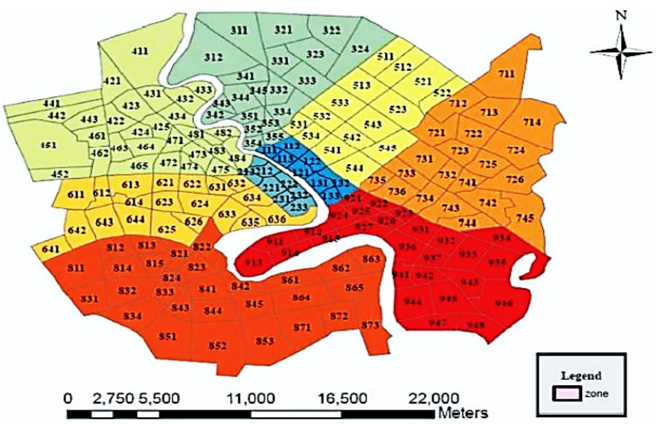

Fig. 1. TAZs of Baghdad city.

D. Household Survey Types

In order to understand the household travel behavior the survey systematically collected detailed trip patterns, like origins, destinations, modes, and purposes, and key household characteristics affecting the mobility, such as income, vehicle ownership, and family structure [21]. Two survey methods were used to collect these data:

- Full Interview Technique: Conducted in-person with all household members, ensuring higher data accuracy and immediate verification. However, this method is resource-intensive.
- Self-Completed Questionnaire Technique: In this method, an interviewer collects household demographic data, while a travel diary is left with the household for self-completion over 1–2 days. This approach is cost-effective but may result in lower response accuracy.

The method used in each zone depended on the number of households and resource availability (Table I).

TABLE I. METHOD DISTRIBUTION

No. of zone	Full interview		Questionnaire		Total	
	No.	%	No.	%	No.	%
901	15	17.6	70	82.4	85	100
903	45	25.9	129	74.1	174	100
905	25	18.9	107	81.1	132	100
907	20	18.7	87	81.3	107	100
909	50	25.4	147	74.6	197	100
911	12	15.0	68	85.0	80	100
913	5	20.0	20	80.0	25	100
915	25	18.8	108	81.2	133	100
923	10	16.4	51	83.6	61	100
925	15	16.7	75	83.3	90	100
929	15	17.4	71	82.6	86	100
Total	237	20.3	933	79.7	1170	100

E. MLR Model Development

Trip production models were developed using SPSS v25 software. The stepwise regression method was selected to construct simplified forecasting models by progressively including statistically significant variables [6, 12, 15]. The modeling procedure involved the following steps:

- Initial Variable Entry: The independent variable with the highest F-test significance entered the model first.
- Iterative Model Building: Additional variables were included if they met the F-test threshold ($F \geq 3.48$) and demonstrated statistical significance ($p < 0.05$).
- Termination Criteria: The modeling process ended when no additional variables met the entry criteria or when further inclusion failed to improve the model fit [21].

This approach enabled the identification of key socioeconomic and demographic predictors influencing the trip generation, thereby contributing to the the development of more accurate travel demand forecasting models.

IV. RESULTS AND DISCUSSION

A. Preliminary Analysis

Before constructing the trip generation models, a preliminary analysis of the collected data was conducted to gain insights that would support the model development.

The gender distribution across the zones shows that the females outnumber the males in TAZs 901, 903, 907, 909, 911, 915, 923, 925, and 929, while the males slightly outnumber the females in zones 905 and 913 (Figure 2).

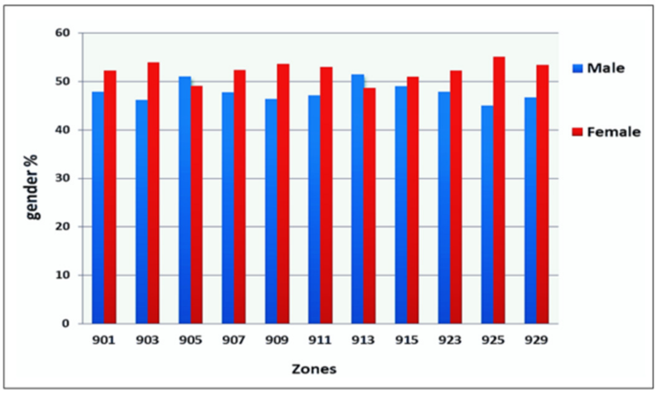

Fig. 2. Gender distribution in TAZs.

The respondents were divided into five groups (Figure 3). The age categories 25–60 years and 6–18 years are the most prominent across all zones, reflecting the dominance of working-age adults and school-aged children. This demographic structure correlates with increased work and educational trip rates.

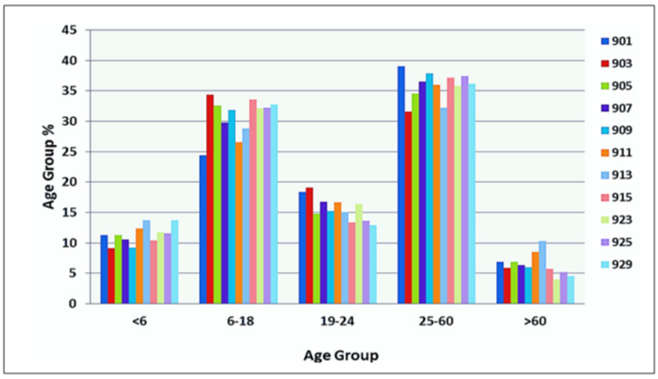

Fig. 3. Age group distribution in TAZs.

Work trips rank second in frequency after educational ones. They are followed by shopping and other types of trips (Figure 4), indicating the importance of non-work travels in daily patterns. From the dominant modes of travel, the private vehicles and buses are the most used (Figure 5), largely due to the average income levels in the study area. The household income was classified into three categories: low, medium, and high, with the medium being the most prevalent across all zones, (Figure 6).

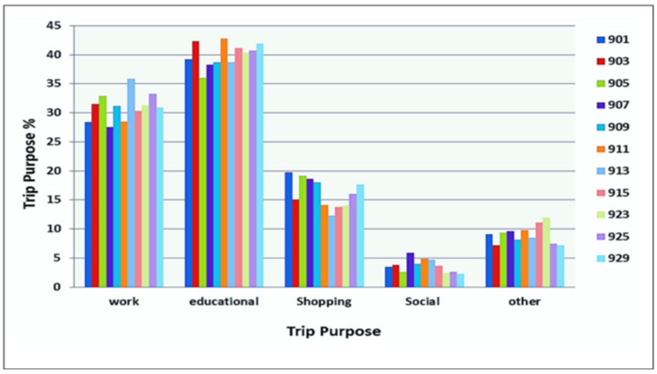

Fig. 4. Types of trip purposes in TAZs.

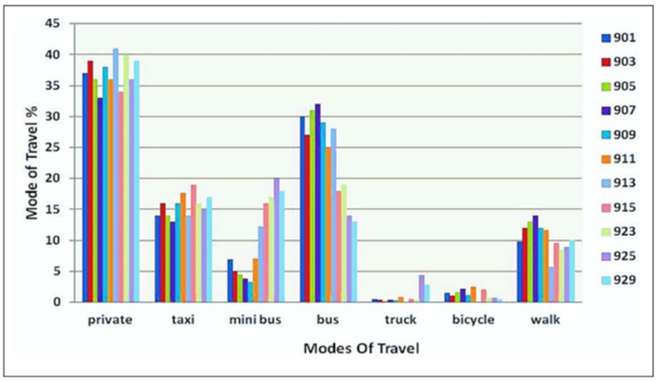

Fig. 5. Travel modes in TAZs.

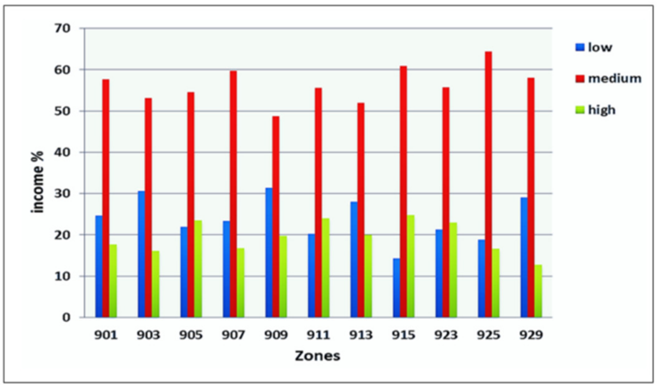

Fig. 6. Distribution of revenue in TAZs.

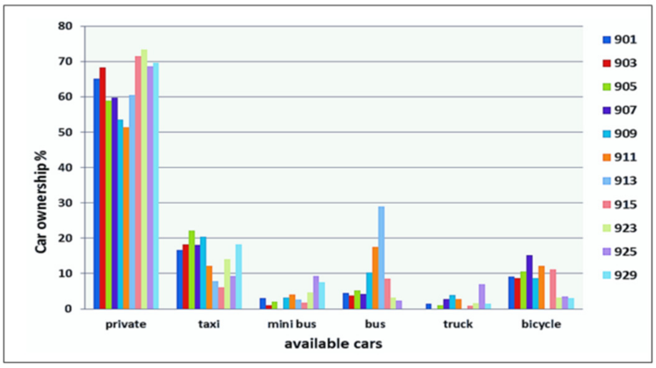

Fig. 7. Car ownership.

Figure 7 presents the car ownership rates across the zones. The high percentage of private vehicle ownership indicates a relatively strong financial capacity, contributing to a higher number of daily trips.

Another important finding is that the dwelling characteristics, including type, ownership, and area, are associated with the average household income and influence the travel behavior (Figure 8).

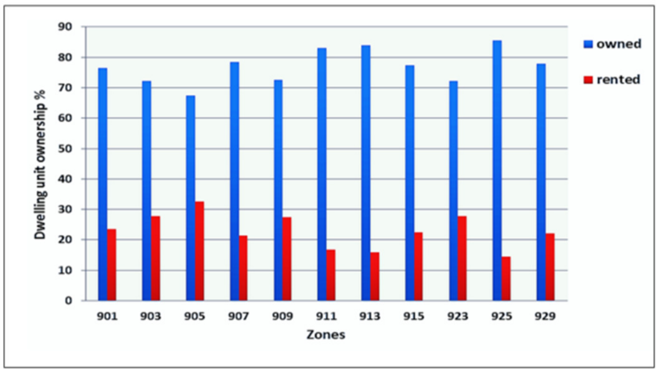

Fig. 8. Dwelling unit ownership in all TAZs.

B. Statistical Analysis

Table II lists the autonomous variables used in the trip production models.

TABLE II. AUTONOMOUS VARIABLES

Item	Variable
X1	Family size
X2	Number of males
X3	Number of females
X4	Number of workers
X5	Number of students
X6	Persons aged less than 6 years
X7	Persons aged 6-18 years
X8	Persons aged 19-24 years
X9	Persons aged 25-60 years
X10	Persons aged more than 60
X11	Household income
X12	Dwelling unit type
X13	Dwelling unit ownership
X14	Dwelling unit area
X15	Number of car owners

To meet the assumptions of MLR, a correlation matrix (Table III) was developed using SPSS v25, to ensure that the independent variables are not significantly collinear [8, 14, 15, 16, 21, 23]. This step validates the suitability of the data for regression modeling.

The descriptive statistics for these variables are presented in Table IV, where the mean and standard deviation values are calculated based on 790 valid observations.

TABLE III. CORRELATION MATRIX FOR THE INDEPENDENT VARIABLES

Variables	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
X1	1.00														
X2	.586	1.00													
X3	.682	.068	1.00												
X4	.523	.277	.167	1.00											
X5	.757	.471	.558	-.060	1.00										
X6	.327	-.013	.231	.080	-.025	1.00									
X7	.639	.410	.458	.075	.423	-.007	1.00								
X8	.341	.191	.271	-.041	.295	.067	.042	1.00							
X9	.438	.311	.232	.529	.057	.093	.099	-.058	1.00						
X10	.174	.189	.054	.097	.019	.100	.054	.055	.114	1.00					
X11	.109	.165	.020	.457	.021	.018	.026	.115	.360	.010	1.00				
X12	-.004	-.003	-.008	-.097	-.046	-.044	.023	-.049	-.082	-.055	-.053	1.00			
X13	-.025	-.031	-.069	-.156	-.101	-.023	-.067	-.072	-.124	.012	-.088	.239	1.00		
X14	.204	.230	.046	.290	.087	-.010	.073	-.014	.240	.032	.139	-.140	-.234	1.00	
X15	.423	.125	-.012	.332	-.049	.014	.060	.023	.512	.060	.239	-.014	-.038	.141	1.00

TABLE IV. DESCRIPTIVE STATISTICS

Item	Mean	Std. Deviation	N
Y	3.8557	1.36382	790
X1	4.8139	1.59992	790
X2	2.2886	1.17332	790
X3	2.5278	1.17800	790
X4	1.6127	0.66743	790
X5	2.1987	1.28153	790
X6	0.4810	0.70595	790
X7	1.4797	1.15425	790
X8	0.8089	0.76201	790
X9	1.7620	0.68781	790
X10	0.2823	0.56296	790
X11	2.0354	0.68942	790
X12	1.1494	0.35668	790
X13	1.2810	0.44978	790
X14	1.4342	0.53529	790
X15	0.7595	0.54494	790

C. Model Validation

Comparing the observed versus the predicted trip generation values for all trips, work-related trips, and educational trips, respectively, validates the model's ability to accurately predict the travel behavior.

The analysis (Figures 9-11) indicates that the most influential variables for work trip generation are: the number of workers in the household (X4), number of car owners (X15), persons aged 25–60 (X9), dwelling unit type (X12), and family size (X1), while for the educational trip generation: the number of students (X5), persons aged 6–18 (X7), persons aged 19–24 (X8), and dwelling unit ownership (X13).

It is worth mentioning that the family size (X1) has a negative effect on the creation of educational trips.

Table V presents the final stepwise regression models developed at a 95% confidence level.

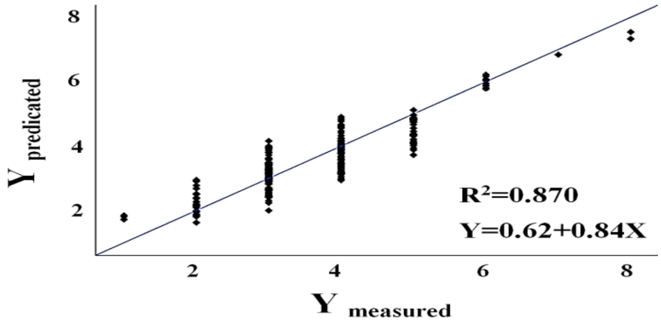

Fig. 9. Measured rate versus the predicated rate of total trip.

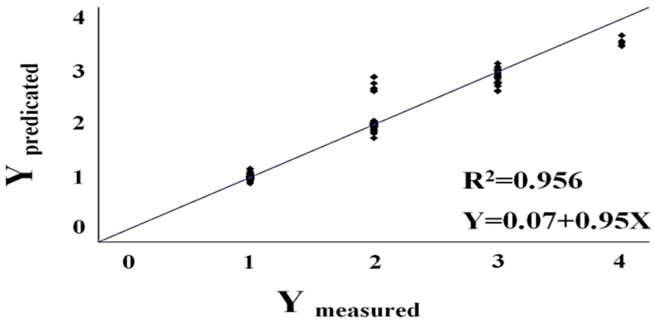

Fig. 10. Measured rate versus the predicated rate of work trip.

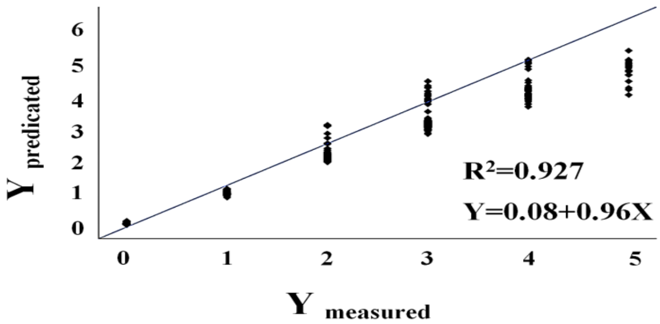

Fig. 11. Measured rate versus the predicated rate of educational trip.

TABLE V. DEVELOPED MODELS

Trip type	Model	R^2	Adj. R^2	S.E.E
Total trips (Y)	0.349+0.678X5+0.797X4+ 0.203X15+0.097X1+0.089X11-0.078X7+0.065X2	0.870	0.862	0.538
Work trip (Y1)	0.076+0.958X4+0.054X15+ 0.031X9+0.027X12+0.019X1	0.956	0.951	0.134
Educational trip (Y2)	0.145+0.860X5+0.087X7-0.057X1+0.064X13+0.051X8	0.927	0.933	0.312

V. CONCLUSIONS

This study analyzed the trip generation patterns in selected zones of Baghdad using Multiple Linear Regression (MLR) modeling. Within the scope of the collected household survey data, the following conclusions were drawn:

The coefficient of determination (R^2) for the trip production model was found to be 0.870, indicating a strong explanatory power of the selected independent variables.

The model's performance was evaluated using standard statistical metrics:

- Mean Squared Error (MSE): 1.27
- Root Mean Squared Error (RMSE): 1.13
- Mean Absolute Percentage Error (MAPE): 27.54%
- Average model accuracy: 72.5%

A high R^2 combined with low MSE and RMSE values indicates that the developed MLR model performs well in forecasting the trip generation. The obtained error metrics also demonstrate improved prediction accuracy compared to similar models in previous studies [3].

The key variables influencing the trip generation were identified as: the number of students in the household, number of workers in the household, family size, gender composition (males, females), household age distribution, household income, and number of car owners.

A total of 5,529 trips were recorded in the study area. Among these, home-based educational trips represented the largest share, accounting for 2,210 trips, strongly associated with the number of students and individuals aged 6–18 years. Work-related trips followed, totaling 1,705 trips, primarily linked to the number of employed individuals and the adult age group (25–60 years).

Moreover, areas with diverse land use and higher service concentration tend to generate a significantly greater number of trips, underscoring the relationship between the urban form and mobility patterns.

These findings highlight the importance of the demographic and socioeconomic characteristics in shaping the urban mobility. The developed models can serve as effective tools for transportation planners, enabling a more accurate forecasting of the travel demand and informed decision-making for infrastructure development in rapidly urbanizing environments. To enhance the utility of trip generation models and improve the transport planning in Baghdad, it is proposed to develop a GIS-based zoning system and road network database.

ACKNOWLEDGMENT

The authors are grateful to the College of Engineering, Al-Iraqia University, for its support in accomplishing this work.

REFERENCES

[1] J. C. Yu, *Transportation Engineering: Introduction to Planning, Design, and Operations*. New York: Elsevier Science Ltd, 1982.

[2] N. H. Majeed and G. J. Qasim, "Trip attraction model of selected zones in Baghdad city," *Journal of Physics: Conference Series*, vol. 1973, no. 1, Dec. 2021, Art. no. 012235, https://doi.org/10.1088/1742-6596/1973/1/012235.

[3] Z. A. A. Salim and G. J. Qasim, "Trip Production Modeling of Selected Zones in Baghdad City," *Journal of Engineering and Sustainable Development*, vol. 24, pp. 126–136, Aug. 2020, https://doi.org/10.31272/jeasd.conf.1.14.

[4] N. M. Asmael, "A GIS-Assisted Optimal Route Selection Based on Transportation Network Design (Baghdad Metro Case Study)," Ph.D. Thesis, Baghdad University, 2015.

[5] A. Jayasinghe, K. Sano, and K. Rattanaporn, "Application for developing countries: Estimating trip attraction in urban zones based on centrality," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 4, no. 5, pp. 464–476, Oct. 2017, https://doi.org/10.1016/j.jtte.2017.05.011.

[6] R. A. Johnson and G. K. Bhattacharyya, *Statistics: Principles and Methods*. John Wiley & Sons, 2019.

[7] M. Amaya, R. Cruzat, and M. A. Munizaga, "Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis," *Journal of Transport Geography*, vol. 66, pp. 330–339, Jan. 2018, https://doi.org/10.1016/j.jtrangeo.2017.10.017.

[8] H. W. Lo, C. J. Pai, and M. Deveci, "A multi-objective model for integrated supplier order allocation and supply chain network transportation planning decision-making," *Information Sciences*, vol. 689, Jan. 2025, Art. no. 121487, https://doi.org/10.1016/j.ins.2024.121487.

[9] M. Kumar and A. K. P., "An innovative modelling framework for freight consolidation in transportation planning," *IIMB Management Review*, vol. 37, no. 1, Mar. 2025, Art. no. 100550, https://doi.org/10.1016/j.iimb.2025.100550.

[10] P. Barman and L. Dutta, "Charging infrastructure planning for transportation electrification in India: A review," *Renewable and Sustainable Energy Reviews*, vol. 192, Mar. 2024, Art. no. 114265, https://doi.org/10.1016/j.rser.2023.114265.

[11] K. M. Almatar, "Investigating the New Public Transportation System in Riyadh City," *Journal of Urban Planning and Development*, vol. 150, no. 1, Mar. 2024, Art. no. 05023051, https://doi.org/10.1061/JUPDDM.UPENG-4331.

[12] K. M. Almatar, "Towards sustainable green mobility in the future of Saudi Arabia cities: Implication for reducing carbon emissions and increasing renewable energy capacity," *Heliyon*, vol. 9, no. 3, Mar. 2023, Art. no. e13977, https://doi.org/10.1016/j.heliyon.2023.e13977.

[13] K. M. Almatar, "Transit-Oriented Development in Saudi Arabia: Riyadh as a Case Study," *Sustainability*, vol. 14, no. 23, Jan. 2022, Art. no. 16129, https://doi.org/10.3390/su142316129.

[14] L. R. Kadiyali, *Transportation Engineering*. Khanna Publishing, 2016.

[15] D. Cramer and D. L. Howitt, *The SAGE Dictionary of Statistics: A Practical Resource for Students in the Social Sciences.* London Thousand Oaks: SAGE Publications Ltd, 2014.

[16] N. Garber and L. Hoel, *Traffic and Highway Engineering*. Stamford, CT, USA: Cengage Learning, 2014.

[17] A. K. Bala and A. Kumar, "Research Progress Knowledge Mapping on Transit Oriented Development (TOD) Typology," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21847–21853, Apr. 2025, https://doi.org/10.48084/etasr.10097.

[18] A. A. Abdulhussein and A. K. Jameel, "Involving Infrastructure as a Latent Variable in Active Transportation Mode Choice: The Case Study of Baghdad City," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18585–18591, Dec. 2024, https://doi.org/10.48084/etasr.8616.

[19] K. M. Almatar, "Smart transportation planning and its challenges in the Kingdom of Saudi Arabia," *Sustainable Futures*, vol. 8, Dec. 2024, Art. no. 100238, https://doi.org/10.1016/j.sftr.2024.100238.

[20] K. M. Almatar, "Increasing electric vehicles infrastructure in urban areas for efficiently employing renewable energy," *Environment, Development and Sustainability*, vol. 26, no. 10, pp. 26183–26204, Oct. 2024, https://doi.org/10.1007/s10668-023-03723-x.

[21] I. of T. Engineers, *Transportation Planning Handbook*, 2nd ed. Washington, DC: Inst of Transportation Engrs, 1999.

[22] J. Guo, G. Trinidad, and N. Smith, "MOZART: A multi-objective zoning and aggregation tool," in *Proceedings of the Philippine Computing Science Congress (PCSC)*, 2000, pp. 197–201.

[23] M. J. Bruton, Introduction to Transportation Planning. UCL Press, 1992.